Recommender System-Based Diffusion Inferring for Open Social Networks

著者	YANG Xiao, DONG Mianxiong, CHEN Xiuzhen, OTA Kaoru
journal or publication title	IEEE Transactions on Computational Social Systems
volume	7
number	1
page range	24-34
year	2020
URL	http://hdl.handle.net/10258/00010328

doi: info:doi/10.1109/TCSS.2019.2950139

Recommender System-Based Diffusion Inferring for Open Social Networks

Xiao Yang, *Student Member, IEEE,* Mianxiong Dong, *Member, IEEE,* Xiuzhen Chen, *Member, IEEE,* and Kaoru Ota, *Member, IEEE*

***Abstract*—Open social network (OSN) plays a more significant role in information propagation through the rapid developing of information technology. Since information diffusion is an essential process happens in OSN, it has been studied in many researches. Several models have been proposed to infer the diffusion process and reproduce diffusion network. However, these methods have two critical problems, 1) ignoring the effects of user social characteristics, 2) inaccuracy resulted from calculating the influence of different features independently. To address these limitations, a diffusion inferring method based on recommender system (DIM-SPTF) was proposed. DIM-SPTF method considers the propagation process between users as the recommendation process of information and employs recommender system to infer the propagation relationship. Through determining the propagation relations among all users in the observed topic data set, information diffusion network can be finally obtained. Experimental results show that DIM-SPTF leads to improvements in performance compared with state-of-the-art methods.**

***Index Terms*—Open social network, diffusion inferring, information propagation, recommender system.**

I. Introduction

OPEN social network (OSN), or what we sometimes call online social network, is an online open platform which people use to build social networks or social relationships and can exert an enormous influence on society through the rapid diffusion of information. Learning the information diffusion process in OSN could find hidden relations between users and dig valuable characteristics of network, which is widely applied in monitoring net environment, network situation analysis, and community detection.

During a diffusion process, several pieces of innovative information diffuse from original users to other users over the whole network like an epidemic. Infected users (users who have received the information) can also infect (spread information) their neighbors with a certain probability and one user may be infected by not only a single user. Diffusion ends when all nodes are infected or when the topic is not transmitted by anyone [1], [2], [3], [4], [5], [6].

Manuscript received XX XX, XX

Xiao Yang and Xiuzhen Chen are with Shanghai Jiao Tong University, 800 Dongchuan Road, Minhang District, Shanghai, China (e-amil: youngshall@sjtu.edu.cn; chenxz@sjtu.edu.cn).

Mianxiong Dong and Kaoru Ota are with Muroran Institute of Technology, 27-1 Mizumoto-cho, Muroran-shi, Hokkaido, Japan (e-amil: mx.dong@csse.muroran-it.ac.jp; ota@csse.muroran-it.ac.jp).

This research was supported by JSPS KAKENHI (Grant No. JP16K00117 and JP19K20250), KDDI Foundation, National Key R&D Program of China (Grant No. 2016YFB0801003) and the National Natural Science Foundation of China (Grant No. 61562004 and No. 61431008). Mianxiong Dong is the corresponding author.

In many cases, it is intractable to directly observe the information diffusion process in OSN because of the privacy protections and the lack of relevant information in the online community [7], [8], [9], [10], [11], [12]. However, observing the corresponding changing characteristics of user nodes in diffusion are relatively simple, because the diffusion process will finally leave some observable data such as tweets, reposted blogs, and comments, which will contribute to the reconstruction of diffusion network. So, this study focuses on utilizing these observed information data in OSN to infer the whole diffusion network through analyzing connectivity between pair users.

The main challenges in inferring diffusion network come from two parts. First, the social characteristics of users are not taken into account when judging the propagation relationship in the diffusion process. In OSN, users may have various social preferences and there may be interactions between users with similar preferences, which will have an impact on information propagation. Second, many features interact rather than independently perform in the whole diffusion process. Accordingly, when inferring diffusion network, the intrinsic connection and the influence between diverse features should be involved.

To tackle these challenges above, this study introduces the concepts of recommender system into inferring diffusion network. Giving some inquiry input e.g. user and contextual information, the recommender system can output some items that is most relevant to the input. System can learn the features of current items and exploit the connections available in existing data to recommend items. For instance, when a user browses a commodity in website, the system will recommend to the user in terms of the features of the commodity and the correlation between users' preferences obtained from purchase record. Recommended commodity is also related to the connection between the user's preferences and the features of commodity. If the user is more interested in the commodity, the system will have a greater possibility to recommend [13], [14], [15].

Our innovation to apply recommender systems to inferring diffusion network is there are a few certain similarities between the two. In OSN, users can be divided into many social groups and users in the same group tend to spread and share information in some specific aspects, for instance, students and teachers from a university can form a group and the topics they discuss and spread tend to be educational or technological contents. On the basis of this, many concepts in recommender system can be utilized to tackle the inferring of diffusion network. Two users in OSN can be regarded as two users in the recommender system, and users' historical information can be viewed as the purchase records in recommender system. The information that diffuses with textual features can be considered as commodities with different features, and then, judging whether there is a diffusion relationship between users is transformed into judging whether system recommend a commodity (information) that a user browsed to another user. When judging diffusion, time is also a prominent parameter, because users tend to view the latest news and therefore short time interval may lead to greater diffusion possibility.

Derived from the analysis above, a diffusion network inferring method based on recommender system (DIM-SPTF) was proposed in this study to address challenges mentioned before. DIM-SPTF is a method that exploits diffusion relationship between all pair users in observed topic data set and thus to infer the whole diffusion network. When deciding diffusion relationship between pair users, users' preferences in OSN community, information text features, and temporal parameter will be input onto the recommender system to determine diffusion relationship between pair users. Preferences are extracted from user's historical data, while the text features are calculated from the information user diffused. The framework of DIM-SPTF is shown in Figure 1.

DIM-SPTF not only considers textual features but also the influences of user social preferences when determining diffusion relations, which will reduce the inaccuracy resulted from time difference and deficiency of enough features. Furthermore, DIM-SPTF takes into account the correlation between users' preferences and information text features rather than independently considering the impact of each parameter, which will facilitate the utilization of the potential connections between various features. Finally, when determining the diffusion relationship of pair users, DIM-SPTF only needs to analyze the features of current user pair, which is for the purpose of avoiding the uncertainty and redundancy for extracting whole features of all users in OSN.

The main contributions of the study can be summarized as follows:

1) To the best of our current knowledge, this study is the first work which introduces conceptions in recommender systems into the information diffusion process and inferring corresponding network.
2) In light of users' preferences and the text features of the propagated information, an improved cascade model was proposed to the express observed data more detailed. A preference calculation method and a text feature extraction method are also be presented to calculate parameters in cascade.
3) Proposed method was evaluated in 3 different types of data sets and the performances showed that proposed method improved on multiple metrics compared with state-of-the-art algorithms, which also suggests the importance of considering user social features when inferring diffusion network.

The remainder of the article is organized as follows. Previous researches on inferring information diffusion network will be discussed in Section 2. The problem of inferring diffusion network is formulated in Section 3. Section 4 introduces how to infer diffusion network by proposed DIM-SPTF method. The experimental are shown in Section 5. Finally, the conclusion of this study will be mentioned in Section 6.

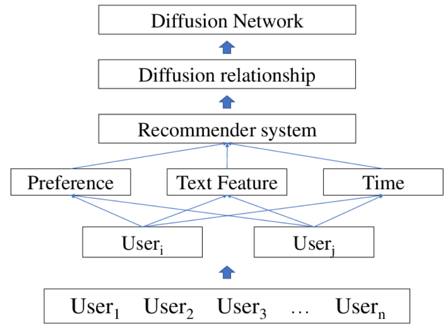


Fig. 1: Framework of diffusion inference by recommender system

II. Related Work

Existing studies in inferring diffusion network can be totally categorized into 2 classes: (i) diffusion cascade based methods, (ii) non-cascade based methods.

For cascade based methods, [16], [17], [18], [19], [20] introduce cascade based method to solve diffusion inferring. Cascade is an observed information sequence from OSN with timestamp and user ID. Cascade based method simplifies the diffusion process and calculates the diffusion probability between pair users in cascade in term of time difference from observed cascade. Subsequently, the mission is to seek a network structure that maximizes the diffusion probability function and the sought network will be selected as the inferred diffusion network [16]. Recently, several models have been proposed to adapt cascade model to practical network situations. In [16], only time difference was calculated in diffusion probability and it's apparently not comprehensive. Hence, [21], [22], [23] utilize more text features to help improve the inferring accuracy in diffusion probability. [21] defines several new probabilistic models which could describe recurrent cascades, and take into account not only the propagation time differences between topic information, but also a greater set of attached features such as language and content to determine diffusion probability. [22] evaluates the propagation attributes of multi-information in social media and presents an Expectation Maximization algorithm to infer the diffusion process with multi-information patterns. In study [23], text features and timestamp are utilized to judge the propagation nodes at each diffusion time steps, and finally construct the whole propagation network. Cascade based methods assume the underlying network to be static, but the network situation is usually dynamic in reality. To cope with this shortcoming, [24] proposes an algorithm that constructs information diffusion network while the diffusion process is both dynamic and potential. Proposed method in [24] models the existence of all edges as a stochastic propagation process and use hidden Markov model to reproduce diffusion process. [25] supposes that users' interactions show various features and propagation speed, and construct a mixed diffusion pattern model to infer the information network of heterogeneity.

Meanwhile, there are non-cascade based methods. Considering the heterogeneity of network structure, [26] proposes NI algorithm to combine with network structure features to infer the diffusion network under heterogeneous networks. Study [27] makes a few improvements on the cascade based method and proposes a trace based method. Different from cascade methods, [27] extracts the state trace of each user at every time step when observing test data, which will be then applied to minimization of the risk function to infer the hidden diffusion network. [28] employs users' feature to build feature tree, and subsequently compares the differences between users and feature tree to judge the user diffusion relationship. Accurate propagation time is difficult to obtain occasionally, hence, [29] applies Bayesian model to calculate the probability of the whole diffusion network without employing temporal parameter to handle this disadvantage. Collective graphical model with noisy aggregate observation was investigated in [30] and a MAP method was presented to promote the propagation probability. [31] raised the problem of identifying information from underlying network and external network. Based on this problem, [31] developed a parametric fitting technology to distinguish and describe the characteristics of information features within internal network and external information. Then these features will be utilized for underlying network construction.

Compared with previous studies, besides user behavioral characteristics and information textual characteristics, DIM-SPTF also involved social features and the feature correlations between users to infer diffusion network, which will be more conducive to improving the accuracy of network construction. Additionally, dissimilar to statistical or epidemiological models, recommender system was exploited to judge the diffusion relationship of users.

III. Problem Statement

Consider an unknown hidden information diffusion network $G^* = (V, E)$ with $V = (u_1, u_2, ...u_n)$ represents all individual user nodes that propagates some topic information and $E = (e_{ij}|i, j \in V)$ represents the diffusion relationship that i transmits information to j. As mentioned in Section 1, to observe the diffusion process directly of a certain topic in OSN is arduous, but with the observed users' data e.g. tweets and blogs left in the diffusion network, it could be feasible to infer the diffusion network. Observed user "infection" (diffusion) data is generally described by cascades. A cascade is a sequence that records the diffusion events

$$c = \{(t_0, u_0), (t_1, u_1), ..., (t_n, u_n)\}$$

Where $u_i \in V$ is a user that receives the topic information at time step t_i, Supposing the diffusion interval of the whole network is T. Note that a cascade is just the observation sequence of infected users without details how users are infected and the problem can be explained as inferring diffusion network G^*, giving corresponding diffusion cascades of the underlying network [16].

IV. Proposed Method: DIM-SPTF

A. Feature Extraction

As referred in Section 3, cascades are observed user data sequence extracted from OSN topics and utilized to infer underlying diffusion network structure. However, it is obviously not appropriate and accurate enough to apply timestamp t_i and user ID u_i in cascades to infer hidden structures, because the social attributes among users will have an impact on the diffusion of topic information, and apparently there is also a certain connection between the characteristics of propagated texts among users. But the relationships between users and the links between propagated texts are not taken into account in previous studies and more relevant information is required. In order to handle this problem, this study will optimize cascades and extract the users' social preferences and text features of propagated information to improve cascades, so as to facilitate the inferring of diffusion relations.

Motivated by the above assumptions, this study reconsiders the user and information textual factors and redefines the required observed cascades from a certain topic diffusion network $G^\varepsilon = (V^\varepsilon, E^\varepsilon)$ in time interval T as follow

$$c = \{(t_0, u_0, p_0, tf_0), (t_1, u_1, p_1, tf_1), ..., (t_m, u_m, p_m, tf_m)\}$$

Where t_i is timestamp when a user issued information that is related to some topic in OSN and $t_i < T$; u_i is the user ID who propagates the topic information; p_i represents the user u_i social preferences calculated from u_i historical data; tf_i means the text features drawn from topic information propagated by users u_i.

Generally, different cascade c observed at different time intervals in OSN will probably be diverse. Therefore, the whole cascade C of the underlying network can be the combination of all observed cascade c.

$$C = (c_1, c_2, ..., c_n) \tag{1}$$

When observing cascades in OSN, if a user issues a piece of information about a certain topic, the corresponding issued information can be extracted. The extraction of issuing time t_i and the user ID u_i is unproblematic since OSN generally has a searching function and by searching the corresponding topic, topic information's corresponding t_i and u_i can be easily obtained e.g. searching Liverpool won the champions league in twitter and extracting users who discuss the topic. But to get the corresponding user preferences p_i and text features tf_i of the propagated information is intractable and it will be discussed in the rest of the section.

1) User Preference p_i Calculation: User preference p_i is a prominent observation parameter in the cascade c. In general, there are several preferences for a certain user in OSN. If user browses a message and the message matches user preferences, users will be interested in them and have a high probability of spreading the message. Meanwhile, many users with similar preferences tend to form a group, which leads to that users in the group tend to share and propagate the information related to their interests, and it usually spreads faster and more widely.

Users' preferences will be reflected in their behaviors and historical information. For example, a football club fan will be more likely to forward, issue or like news and information related to sports. Hence, users' preferences can be acquired by evaluating their behaviors and corresponding historical information. Moreover, User preferences in general fields will not change much in short time, so, to simplify the study, it's assumed the user preferences are static during the diffusion process.

In this study, users' preferences are calculated by considering the mapping matrices of users to behaviors and behaviors to preferences.

The user-to-behavior mapping is

$$MappingA(u_i) = \{a_j | a_j \in Act, \theta(u_i, a_j) = 1\} \tag{2}$$

Where Act represents the set of possible user behaviors in OSN e.g. issuing, sharing, comment and forwarding, and $\theta(u_i, a_j) = 1$ indicates that user u_i performs a certain behavior a_j.

Suppose there are x users in the cascade c and users have total y kinds of behaviors, then the user-to-behavior matrix is given by

$$B = \begin{bmatrix} b_{11} & b_{12} & ... & b_{1y} \\ b_{21} & b_{22} & ... & b_{2y} \\ ... & ... & ... & ... \\ b_{x1} & b_{x2} & ... & b_{xy} \end{bmatrix} \tag{3}$$

Where $b_{ij} = 1$ indicates that user u_i performs behavior a_j, while $b_{ij} = 0$ is opposite. Each row in Eq.(3) represents behaviors of a user, and each column indicates a certain behavior.

User historical data under certain behavior reflects user preference. Thus the behavior-to-preferences mapping is

$$MappingP(a_j) = \{p_k | p_k \in Pre, \xi(a_j, p_k) = 1\} \tag{4}$$

In Eq.(4), Pre represents the set of user preferences. Through the analysis of huge amount of OSN informations, user preferences Pre could be totally categorized into 9 classes: science, technology, economy, sports, arts, entertainment, social life, politics and military [32]. $\xi(a_j, p_k) = 1$ demonstrates that user data under behavior a_j shows association with preference p_k. Determining the association between behavior and preferences is dependent on the classification of historical information to user preferences. The historical information issued by the user under certain behavior will be classified into one of the 9 preference categories in Pre and semantic classification algorithm in [33] will be employed to do that. Through behavior-to-preferences mapping, the behavior-to-preferences matrix is

$$V = \begin{bmatrix} v_{11} & v_{12} & ... & v_{19} \\ v_{21} & v_{22} & ... & v_{29} \\ ... & ... & ... & ... \\ v_{y1} & v_{y2} & ... & v_{y9} \end{bmatrix} \tag{5}$$

Where $v_{ij} = 1$ reveals that behavior a_i shows association with preference p_j, while $v_{ij} = 0$ is opposite; in Eq.(5), each row illustrates behaviors and column infers preferences.

According to the transfer relation of the mapping function, the preference matrix of cascade c can be calculated by

$$P = B \times V = \begin{bmatrix} b_{11} & b_{12} & ... & b_{1y} \\ b_{21} & b_{22} & ... & b_{2y} \\ ... & ... & ... & ... \\ b_{x1} & b_{x2} & ... & b_{xy} \end{bmatrix} \times \begin{bmatrix} v_{11} & v_{12} & ... & v_{19} \\ v_{21} & v_{22} & ... & v_{29} \\ ... & ... & ... & ... \\ v_{y1} & v_{y2} & ... & v_{y9} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & ... & p_{19} \\ p_{21} & p_{22} & ... & p_{20} \\ ... & ... & ... & ... \\ p_{x1} & p_{x2} & ... & p_{x9} \end{bmatrix} \tag{6}$$

Where each row i of calculated matrix P refers the preferences p_i of each user u_i, which is shown in cascade c.

$$p_i = (p_{i1}, p_{i2}, ..., p_{i9}) \tag{7}$$

2) Text Features tf_i Calculation: Texts with diffusion relationship between two users usually have some similarities in their features. If user forwards a message after reading another message, the feature similarity of the two messages is exactly the same. However, some users in the OSN may not choose to forward a message after reading it, but will probably make some modifications and issue it as original content, or just publish some relevant comments of their own. But even then, new information only changes some expression forms on the basis of the original one, and its core ideas are still obtained from original information. Consequently, it will also contribute to high text features correlation. Therefore, text features should be estimated when reforming the diffusion network of OSN.

When information text is edited by OSN user, different people may have education differences and distinctive views on topic, which make the information text features of same topic vary. With the purpose of making the text features to facilitate analysis, number of available text features should be refined to a certain range.

In this study, connections between preferences p_i and textual features tf_i are mainly considered. Preferences p_i are divided into several common categories, and similarly, textual features tf_i of information are defined as correlation between text content and the 9 preferences in Pre.

$$tf_i = (i_{i1}, i_{i2}, ..., i_{i9}) \tag{8}$$

Where tf_i is the text feature and i_{ij} indicates the relevance between information text and preference j in Pre.

Text content is composed of words, and consequently, the relevance of text content to a certain preferences j is determined by the connections of words in each text to the preference j. The definition of i_{ij} is given by

$$i_{ij} = \sum_k Dis(KeyWord_k, SinPre_j) \times Weight_k \tag{9}$$

Where Dis is the function to compute the semantic distance between words and algorithm in [34] is applied; $KeyWord_k$ means the text keywords obtained by key words extraction technology from the information text propagated by the user; $Weight_k$ is the corresponding weight for $KeyWord_k$; $SinPre_j$ is a certain preference in Pre. Note that not all words in text are included in the calculation of relevance, because some words have little meaning for text analysis, such as stop words and punctuation. Comparatively, keywords can better reflect the essential ideologies of the text. Meanwhile, the results are more reasonable when the correlation is calculated with weighting.

B. Diffusion Relationship Inference

The whole diffusion network structure can be inferred through inferring the propagation relationship between all pair users in observed cascades. In general, DIM-SPTF supposes the information propagated by users in OSN network can be viewed as item, then inferring the diffusion relationship between two users is treated as the process of deciding whether to recommend information to another user by recommender system.

Modern recommender system algorithms are mainly developing from machine learning technology, because the feature representation of users is straightforward and recommendation process is greatly efficient. The general framework of recommender system is composed of 2 main components: wide learning component and deep learning component. Wide learning part is usually generalization model, which holds the ability of capturing the direct user features from observed data and the deep learning part can produce more generalized features and abstract presentations to give precise and rational recommendation.

For this study, user preference and text features are extracted and calculated directly from the observed cascades, so the wide learning component is not necessary and this part will focus on applying the data from the observed cascades to the deep learning component.

DIM-SPTF method adopts recurrent neural network (RNN) as the internal network structure of the recommender system. RNN is highly appropriate for sequential input processing. Hence, it is regarded as the best selection of dealing with dynamic time sequences and sequential observed characteristics data, such as voice signal, image, and text [35].

The input data for recommender system in this method is the cascades of user information data propagated through the underlying OSN. Obviously, these cascades are observed in chronological order. Secondly, there should be characteristic correlations between some adjacent users in cascade, because they all spread the information about the same topic in a very short time interval. Therefore, it is reasonable to adopt RNN as the model to determine diffusion relationships.

1) System Input: System inputs can be extracted from the whole cascades C. Suppose u_1 and u_2 are two users in a cascade c, the user pair information data can be expressed as corresponding input for system

$$input_{12} = (t_1, t_2, p_1, p_2, tf_1, tf_2) \tag{10}$$

Where t_1, t_1 are temporal parameters, p_1, p_2 are user social preferences and tf_1, tf_2 are textual features of propagated informations.

2) System Model: Recommender system in this study employs RNN to learn historical diffusion information to infer propagation relationships in OSN.

RNN is a kind of neural network used to process sequential data, and it includes input layer, hidden layer and output layer. The output in each lay is controlled by output from previous lay, the weights between connected layers and activation function. Unlike feedforward neural network which can only output forward, RNN loops state through the neural network and therefore it is more suitable for time series structures.

The data of the input layer is given by Eq.(10). The output in hidden layer is given by

$$h^{(t)} = tanh(Wh^{(t-1)} + Ux^{(t)} + b_n) \tag{11}$$

$$o^{(t)} = softmax(Vh^{(t)} + b_o) \tag{12}$$

The activation function for the calculation result of input layer to hidden layer is hyperbolic tangent function ($tanh$), and $softmax$ function is employed. According to Eq.(11), the calculation of $h^{(t)}$ would be brought before h into. Consequently, the current output value $h^{(t)}$ of the hidden layer is affected by both the current input $x^{(t)}$ and the all past input $(x^{(t-1)}, x^{(t-2)}, ..., x^{(2)}, x^{(1)})$. Hence, output value $h^{(t)}$ of the hidden layer can be seen as the memory of network, which makes it capable for processing data with temporal correlations.

For the designed model, the loss function should be employed to optimize it. In this study, the cross entropy will be taken to calculate the loss of all samples at a certain time t and the total loss value L is the sum of loss at different times. Loss function equations are shown as follows

$$L^{(t)} = -\frac{1}{m}\sum_{i=1}^{n} y_i^t log(o_i^{(t)}) \tag{13}$$

$$L = \sum_{t=1}^{T}(L^{(t)}) \tag{14}$$

Loss function L is required to be minimized so as to optimize the RNN model, and Adam optimizer is chosen to minimize the loss function. Adam optimizer iteratively upgrades parameter by

$$\begin{cases} m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t & \text{(15a)} \\ v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2 & \text{(15b)} \\ \hat{m}_t = \dfrac{m_t}{1-\beta_1^t} & \text{(15c)} \\ \hat{v}_t = \dfrac{v_t}{1-\beta_2^t} & \text{(15d)} \\ \theta_{t+1} = \theta_t - \dfrac{\eta}{\sqrt{\hat{v}_t + \epsilon}} \hat{m}_t & \text{(15e)} \end{cases}$$

Where θ is the parameter to be upgraded; g_t is the gradient of loss function; m_t, v_t are moment vectors and β_1, β_2 are corresponding exponential decay rates; η is learning rate; ϵ is adjustment parameter.

Recommender system model will be trained to learn the correlations between users' social preferences and information text features before applied to judge the diffusion relationship of network users. The training mode is supervised learning, with training data sets extracted from 9 different types of topics observed from Sina microblog platform and label being the corresponding propagation relationships. The data set information used for training is shown in Table I.

TABLE I: Information about training data sets

Data Set	Amount of User Nodes	Amount of Edges
1	841	1932
2	755	1224
3	912	2072
4	729	1274
5	413	864
6	591	1106
7	921	2160
8	663	1583
9	707	1494

We evaluated the performance of the model with different layers and neurons and found that the training result was best when the hidden layers number was 2 and the neurons was 38. The relevant parameters of the model are revealed in Table II.

TABLE II: Model information

Parameter	Value
Algorithm	RNN
Hidden layer nodes	38
Dropout	0.5
Batch size	32
Number of epoch	80

The trained recommender model is then employed to infer diffusion network. Eq.(10) is used to construct input from observed cascades C and the propagation relationship between pair users will be judged through the recommender system model. Finally, the diffusion network can be inferred by judging the propagation relations among all users in cascades.

V. Experiment

To explore the precise method performance, experiment will test DIM-SPTF under 2 synthetic data sets A, B and real world data sets. Several evaluation parameters in mathematics and machine learning will be utilized to analyze the quality of proposed method. Meanwhile, cascade based DDNE algorithm in [24] and non-cascade based FNI algorithm in [28] will be used as compared baselines. The feature space amounts of DDNE and FNI are 1 and 6-15 respectively, while DIM-SPTF is 10, which shows more stability and efficiency than DDNE and FNI.

A. Synthetic Data Set Generation

Proposed DIM-SPTF algorithm will first evaluated under data sets A. This study applied the model in [36] to simulate the process of information transmission among the users who are extracted from the OSN platform Sina microblogs, so as to generate corresponding observation data sets A. The model simulates the information of users as chromosomes and different information features as genes on chromosomes. When information propagates between two users, information amount of the two users and the density of topic in the network will be utilized to determine whether to propagate. We will generate 6 diffusion networks in various scales for A to test methods performance in different data scales.

The related information for synthetic data sets A is shown in Table III.

TABLE III: Information about data sets A

Data Set	Amount of User Nodes	Amount of Edges
1	100	427
2	500	1678
3	1000	4061
4	5000	6330
5	10000	18141
6	20000	29459

Synthetic data sets B generated by reference [37] will also be applied to evaluate method performance. [37] exploits user tree to build information diffusion network through tree structure and person patterns. We generated 6 data sets in similar scales but add random time delay in each propagation relation to inspect the robustness and accuracy of proposed method.

The related information for synthetic data sets B is shown in Table IV.

TABLE IV: Information about data sets B

Data Set	Amount of User Nodes	Amount of Edges	Time Delay
1	500	1010	0
2	500	964	1 hour
3	500	986	2 hours
4	500	842	3 hours
5	500	975	4 hours
6	500	932	5 hours

B. Real World Data Sets

Besides synthetic data sets, DIM-SPTF will also be tested by real world data sets. In this study, Sina microblog was selected as the object platform to observe the transmission cascades of several hot topics as test data. All the data will be observed from the beginning of spread of topics until the death of the transmission process. To objectively evaluate the performance of the proposed algorithm, we observed 5 data sets with different topics and scales.

The related information for the real world data sets is shown in Table V.

TABLE V: Information about real world data sets

Data Set	Amount of User Nodes	Amount of Edges
1	183	378
2	636	1465
3	1149	3754
4	2328	4559
5	4221	9342

C. Experiment Results

Results will be discussed in this part. First, we will introduce the evaluation standards, and then performance of experiments on both synthetic data sets and real world data sets will be illustrated. Finally, visualization of diffusion network will be analyzed. The platform to running tested methods is Windows 10 with 16GB RAM, Intel Core i7-7700 CPU and Python.

1) Evaluation Standards: In this study, $accuracy$, $precision$, $recall$, $F-score$ are utilized as standards to evaluate the method's performance. Suppose there are two diffusion relation situations among network users: existent propagation relation and non-existent propagation relation and we define 4 parameters: the amount of correctly judged existent propagation relations is TP; the amount of correctly judged non-existent propagation relations is TN; the amount of non-existent propagation relations that are wrongly judged as existent propagation relations is FP; the amount of existent propagation relations that are wrongly judged as non-existent propagation relations is FN. Then the evaluation standards are defined as follows

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- $accuracy$ is the ratio of the correctly judged diffusion relation situations to the whole pool of diffusion relation situations.

$$precision = \frac{TP}{TP + FP}$$

- $precision$ is the ratio of the correctly judged existent propagation relations to all the judged existent propagation relations. We can see that $precision$ talks about how precise the model is out of those predicted existent propagation relations and how many of them are actually existent propagation relations.

$$recall = \frac{TP}{TP + FN}$$

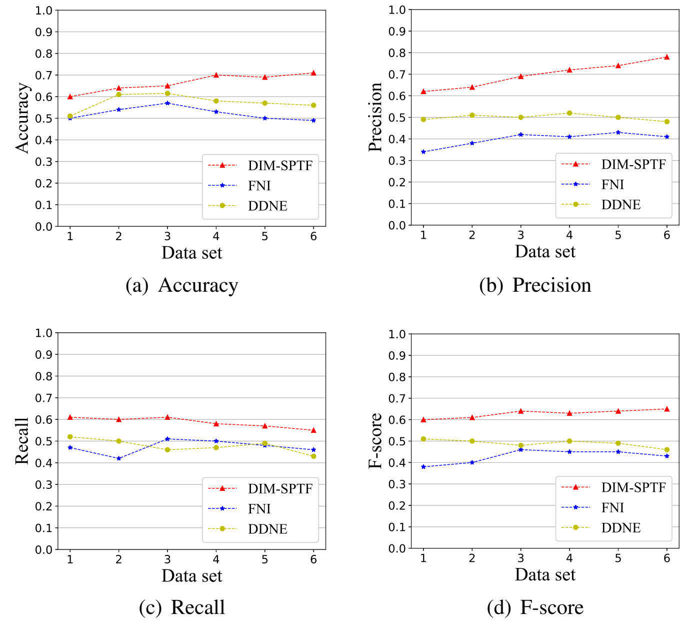
- $recall$ is the ratio of the correctly judged existent propagation relations to all the existent propagation relations in network. So $recall$ actually calculates how many of the actual existent propagation relations that our model capture through labeling it as existent propagation relations from the real existent propagation relations.

$$F - score = 2 \times \frac{recall \times precision}{recall + precision}$$

- $F - score$ is the harmonic mean (average) of the $precision$ and $recall$. Hence $F - score$ will probably be a better measure to use if it's necessary to make a balance between $precision$ and $recall$.

2) Results of Synthetic Data Sets: The results of the experiment on synthetic data sets A are shown in Figure 2. Note that A is utilized to evaluate methods in different data set scales and the data set scale in A increases with the data set number.

As shown in Figure 2(a), with the increase of the scales of data sets, the accuracies of the tested methods have not changed much. The overall $accuracy$ of DIM-SPTF is between 60% and 71% and it's about 10% higher than FNI and 8% higher than DDNE averagely.



(a) Accuracy (b) Precision (c) Recall (d) F-score

Fig. 2: Experiment results on synthetic data sets A

The performances in $precision$ are shown in Figure 2(b). $precision$ of DIM-SPTF rises with the increase of data set scales and the overall $precision$ is about 61% to 78%, while FNI and DDNE are almost stable and keeps $precision$ at about 41% and 50% respectively. DIM-SPTF improves $precision$ by about 24% compared with FNI and is about 15% higher than DDNE in total. This indicates that DIM-SPTF shows better efficiency in finding existent propagation relationships precisely from judged relations, which is probably because DIM-SPTF considers not only more detailed features such as social preferences among users and text, but also the relations between multiple features.

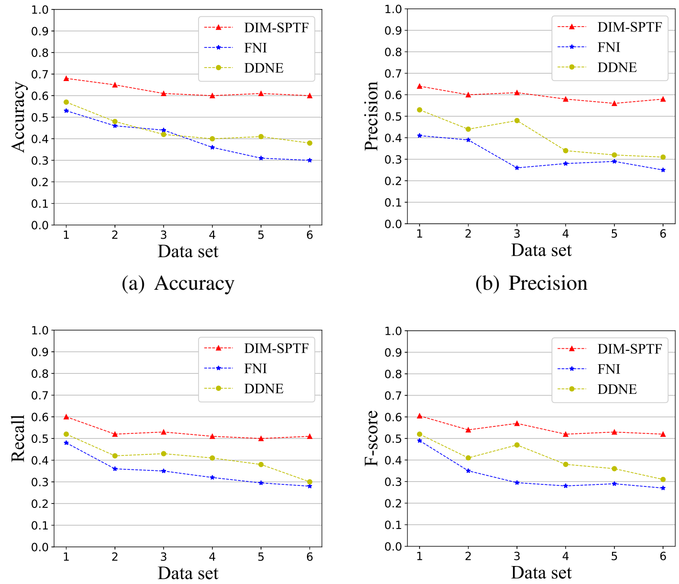
Figure 2(c) shows the performances of $recall$ rates. DIM-SPTF algorithm is about 57%, while FNI fluctuates around 47% and DDNE keeps at about 49%. DIM-SPTF performs better on $recall$ rate, which suggests that DIM-SPTF is more proficient in digging more existing propagation relations from original network.

$F-score$ measures the overall performance, and it's shown in Figure 2(d). For DIM-SPTF, $F - score$ increases slightly with data scales, while FNI and DDNE change around certain values. The average $F - score$ of DIM-SPTF is about 61%, greater than that of FNI (45%) and DDNE (49%).

The results of the experiment on synthetic data sets B are shown in Figure 3. Note that B is exploited to test method robustness under different propagation delay and the delay rises along with the data set number.

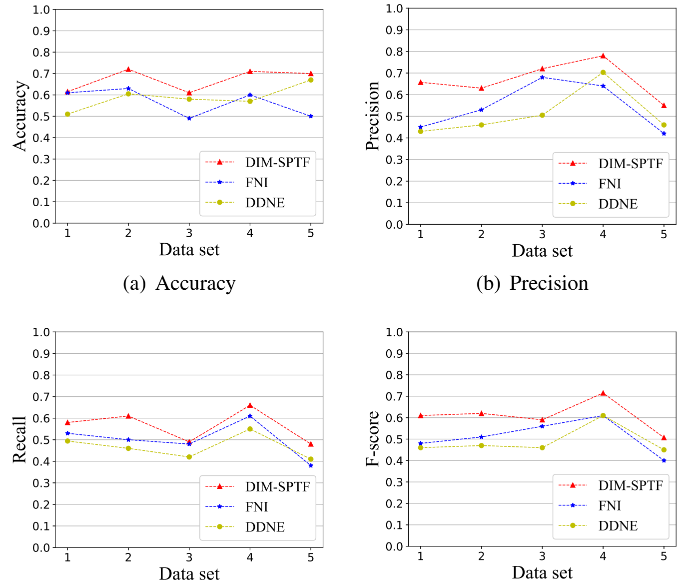
Since delay increases, performances of all the tested methods declined in varying degrees.

In Figure 3(a), with the increase of time delay, the $accuracy$ of DIM-SPTF decreases by about 9%, while FNI and DDNE



(a) Accuracy (b) Precision (c) Recall (d) F-score

Fig. 3: Experiment results on synthetic data sets B



(a) Accuracy (b) Precision (c) Recall (d) F-score

Fig. 4: Experiment results of real world data sets

decrease by 22% and 17% respectively. DIM-SPTF is generally higher than FNI and DDNE algorithm in $accuracy$. This demonstrates that DIM-SPTF has better robustness against time delay than FNI and DDNE, because for DIM-SPTF, time is not the only decisive effect when judging propagation relation.

Figures for $precision$, $recall$, and $F-score$ illustrate alike curve variation characteristics, in which DIM-SPTF and DDNE tend to have slight vibrations, while FNI keeps falling with time delay rising. From the perspective of numerical changes, FNI and DDNE decreases more significantly, while DIM-SPTF outperforms and is relatively stable.

Totally speaking, DIM-SPTF outperforms DDNE and FNI. DDNE take account of temporal parameter and FNI employed user features. Both of DDNE and FNI didn't analyze textual features and social features, which leads to low efficiency in results of A. For data set B, DIM-SPTF considered the features interactions, while DDNE and FNI consider features perform independently, so DIM-SPTF runs more stably.

3) Results of Real World Data Sets: The result of real world data sets are shown in Figure 4. Note that data set scale increases with the data set number.

In terms of $accuracy$ from Figure 4(a), the average $accuracy$ of DIM-SPTF is about 65%, and FNI and DDNE is about 55% and 58%. Although the $accuracy$ fluctuates with the data sets, DIM-SPTF promotes the performance by about 9% totally over the comparison methods.

For $precision$ in Figure 4(b), DIM-SPTF is relatively superior compared with baseline methods, and hold average $precision$ at about 67%. Meanwhile, the precisions of FNI and DDNE is not as high as DIM-SPTF, with average values around 56% and 51%.

In Figure 4(c), $recall$ rates of tested methods have a similar overall trend like the performances in $precision$. But the performance differences between DIM-SPTF and comparison algorithms are relatively minuscule.

The overall performances of $F-score$ are shown in Figure 4(d). With the magnification of data set scales, although the performance difference is not significant, DIM-SPTF still maintains a lead of about 8% performance compared with FNI and DDNE.

For the result of real world data set, DIM-SPTF is better than comparisons like the results of set A. However, FNI outperforms DDNE, this is probably because real user data set shows more user features so that FNI can better describe diffusion model.

4) Visualization of Diffusion Network: To better explore the nature of information diffusion network and check whether the inferred network is reasonable and correspond with common network characteristics, we build the visualized structure of the inferred network and then do analysis. Data set 5 of the real world data sets is selected as the target network, which can better display the diffusion characteristics through its large data scale. Parameter information about data set 5 is shown in Table VI. The virtualized network figure will not illustrate all user nodes, because user amount of diffusion network is too massive, so we choose the influential nodes to research. According to the sum of comment, like and share of user nodes, influential nodes were divided into three categories: the most influential nodes, the average influential nodes, and the less influential nodes and Table VII is the mapping table. The virtualization image is shown in Figure 5.

As shown in Figure 5, the main part of network are the less influential users, while the most influential users and average influential users are small part of the whole. Several small nodes often act as intermediary nodes between influential nodes to help information spread and many nodes tend to cluster together in network, which is consistent with the fact

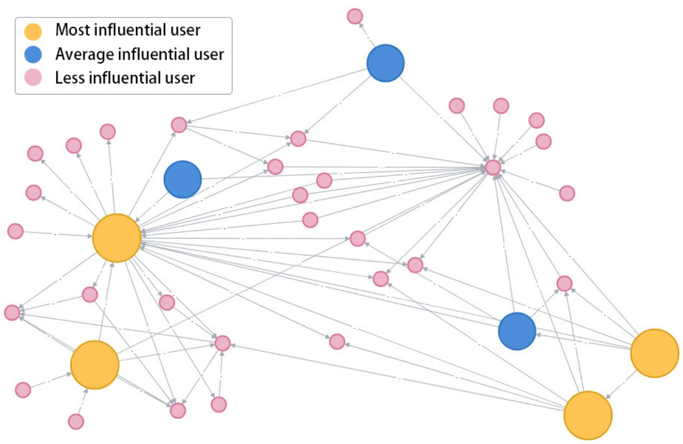


Fig. 5: Inferred diffusion network

TABLE VI: Parameter information of data set 5

Parameter	Value
Topic	USA Huawei ban
User amount	4221
Propagation relations	9342
Most influential nodes	4
Average influential nodes	3
Less influential nodes	31

TABLE VII: Mapping table for user category

Sum of Comment, Like and Share	Category
Sum $\geqslant$ 1000	Most influential node
500 $<$ Sum $<$ 1000	Average influential node
200 $<$ Sum $\leqslant$ 500	Less influential node

that users tend to form groups as mentioned in Section 1. The overall user preferences have been also illustrated in Figure 6 and it can be inferred that users mainly fo focus on politics, technology and economy, which is matches the topic type of "USA Huawei ban".

OSN users are generally classified into official media nodes and self-media nodes. We count the node types in Figure 5 and list the result in Figure 7. It can be inferred that the amount of self-media users accounts for the vast majority of users in network, but the amount of official media users

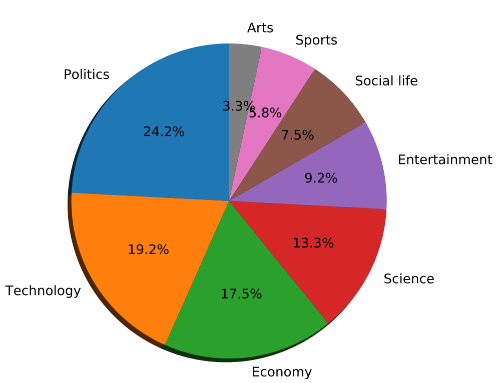


Fig. 6: Overall distribution of user preferences

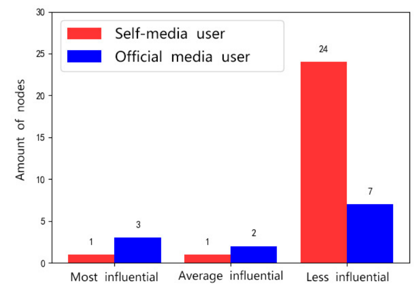


Fig. 7: Amount of influential users

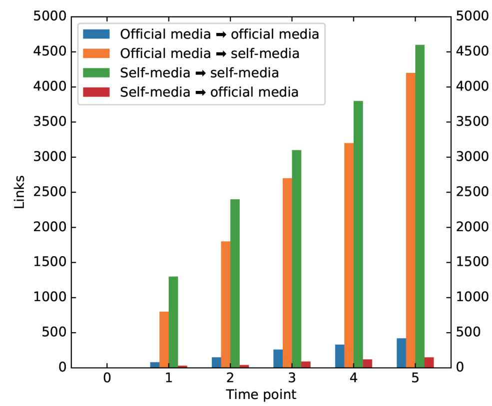


Fig. 8: Propagation link amount between different kinds of user nodes

is relatively large in the most influential nodes and average influential nodes.

We also observe the changes in the amount of node propagation links between different kinds of nodes and the results are shown in Figure 8. Definitely, the amount of all links increases with time, but the amount of links that propagated to self-media users is much greater than others, which is consistent with that self-media users are the main body of diffusion. Moreover, official media users could also influence and spread information to many users, while official media users mostly have their own sources of information, hence, little information is propagated to official media users, which is shown in the blue and red columns.

Figure 9 explores the median propagation time between users. In Figure 9, propagation time transmitted from official media users is mainly distributed in short intervals and it shows that diffusion speed of official media users is relatively fast, which may be due to the quick information acquisition method and reliable accuracy of official media, so users trust official media and would like to diffuse information from it. While the propagation time of self-media users to self-media users distributed more uniformly than others, which may result from that the followers of self-media are relatively less.

In light of the analysis above, inferred information diffusion network conforms to the common characteristics of user amount, network topology, influential nodes and diffusion

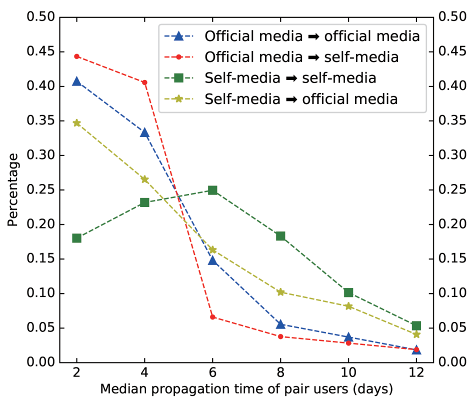


Fig. 9: Median propagation time in network

time.

VI. Conclusion

In this study, we explored the problem of constructing information diffusion network in OSN and proposed a DIM-SPTF method to reproduce information diffusion network and depict relevant propagation characteristics. Firstly, optimizations was made on the original observed cascade model to express users' preferences appropriately and unify the expression of text features. Secondly, to infer the propagation relationship between user pairs in network, several concepts of the recommender system are introduced. Through regarding the process of information diffusion as the recommending of commodity between two users in recommendation, DIM-SPTF adopts the recommender system model to infer the propagation relationships between all pair users in observed cascades, and thus to reconstruct the diffusion network. Finally, proposed DIM-SPTF method was investigated under various scales of synthetic and real world data sets and experimental results reveal that DIM-SPTF method outperforms compared state-of-the-art algorithms in both accuracy and robustness. This study also evaluates the social and structural characteristics of the inferred diffusion network. It is found that in OSN, the majority of influential users are self-media users, who are the main part of diffusion network, while the official media users tend to have great influence and quicker propagation speed in network.

For the future work, we'd like to employ social relationship to diffusion inferring. In OSN, user relation is changeable through time, which makes it intractable to observe and describe. Hence, we are interested in applying topology theory and diffusion models to our method and improving the inferring accuracy.

References

[1] B. Xu and L. Liu, "Information diffusion through online social networks," in *2010 IEEE International Conference on Emergency Management and Management Sciences*, Aug 2010, pp. 53–56.

[2] J. Xu, K. Ota, and M. Dong, "Fast networking for disaster recovery," *IEEE Transactions on Emerging Topics in Computing*, pp. 1–1, 2019.

[3] S. Roy, P. Dey, and D. Kundu, "Social network analysis of cricket community using a composite distributed framework: From implementation viewpoint," *IEEE Transactions on Computational Social Systems*, vol. 5, no. 1, pp. 64–81, March 2018.

[4] D. Vimalajeewa, S. Balasubramaniam, B. O'Brien, C. Kulatunga, and D. P. Berry, "Leveraging social network analysis for characterizing cohesion of human-managed animals," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 2, pp. 323–337, April 2019.

[5] M. Dong, T. Kimata, and K. Zettsu, "Service-controlled networking: Dynamic in-network data fusion for heterogeneous sensor networks," in *2014 IEEE 33rd International Symposium on Reliable Distributed Systems Workshops*, Oct 2014, pp. 94–99.

[6] C. Huang, J. Marshall, D. Wang, and M. Dong, "Towards reliable social sensing in cyber-physical-social systems," in *2016 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, May 2016, pp. 1796–1802.

[7] M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones, "A high-resolution human contact network for infectious disease transmission," *Proceedings of the National Academy of Sciences*, vol. 107, no. 51, pp. 22 020–22 025, 2010.

[8] J. Wu, M. Dong, K. Ota, J. Li, and Z. Guan, "Fcss: Fog computing based content-aware filtering for security services in information centric social networks," *IEEE Transactions on Emerging Topics in Computing*, pp. 1–1, 2019.

[9] D. Perna, R. Interdonato, and A. Tagarelli, "Identifying users with alternate behaviors of lurking and active participation in multilayer social networks," *IEEE Transactions on Computational Social Systems*, vol. 5, no. 1, pp. 46–63, March 2018.

[10] D. Goldenberg, A. Sela, and E. Shmueli, "Timing matters: Influence maximization in social networks through scheduled seeding," *IEEE Transactions on Computational Social Systems*, vol. 5, no. 3, pp. 621–638, Sep. 2018.

[11] X. Liu, D. He, and C. Liu, "Information diffusion nonlinear dynamics modeling and evolution analysis in online social network based on emergency events," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 1, pp. 8–19, Feb 2019.

[12] J. Wu, M. Dong, K. Ota, J. Li, L. Guo, and G. Li, "Chance discovery based security service selection for social p2p based sensor networks," in *2015 IEEE Global Communications Conference (GLOBECOM)*, Dec 2015, pp. 1–6.

[13] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *ACM Computing Surveys*, vol. 52, no. 1, pp. 5:1–5:38, Feb. 2019.

[14] H.-T. Cheng, L. Koc, J. Harmsen, and T. Shaked, "Wide & deep learning for recommender systems," in *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*. New York, NY, USA: ACM, 2016, pp. 7–10.

[15] P. Resnick and H. R. Varian, "Recommender systems," *Communications of the ACM*, vol. 40, no. 3, pp. 56–59, 1997.

[16] M. Gomez-Rodriguez, J. Leskovec, and A. Krause, "Inferring networks of diffusion and influence," *ACM Transactions on Knowledge Discovery from Data*, vol. 5, no. 4, pp. 21:1–21:37, Feb. 2012.

[17] M. Gomez Rodriguez, J. Leskovec, and B. Schölkopf, "Structure and dynamics of information pathways in online media," in *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*. New York, NY, USA: ACM, 2013, pp. 23–32.

[18] S. Myers and J. Leskovec, "On the convexity of latent social network inference," in *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2010, pp. 741–749.

[19] M. G. Rodriguez, J. Leskovec, D. Balduzzi, and B. Scholkopf, "Uncovering the structure and temporal dynamics of information propagation," *Network Science*, vol. 2, no. 1, pp. 26–65, 2014.

[20] M. G. Rodriguez, D. Balduzzi, and B. Schölkopf, "Uncovering the temporal dynamics of diffusion networks," *28th International Conference on Machine Learning (ICML)*, 2011.

[21] L. Wang, S. Ermon, and J. E. Hopcroft, "Feature-enhanced probabilistic models for diffusion network inference," in *Machine Learning and Knowledge Discovery in Databases*. Springer Berlin Heidelberg, 2012, pp. 499–514.

[22] S. Wang, X. Hu, P. S. Yu, and Z. Li, "Mmrate: Inferring multi-aspect diffusion networks with multi-pattern cascades," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, 2014, pp. 1246–1255.

[23] T. M. Snowsill, N. Fyson, T. De Bie, and N. Cristianini, "Refining causality: Who copied from whom?" in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, 2011, pp. 466–474.

[24] M. Tahani, A. M. A. Hemmatyar, H. R. Rabiee, and M. Ramezani, "Inferring dynamic diffusion networks in online media," *ACM Transactions on Knowledge Discovery from Data*, vol. 10, no. 4, pp. 44:1–44:22, Jun. 2016.

[25] C. Su, X. Guan, Y. Du, X. Huang, and M. Zhang, "Toward capturing heterogeneity for inferring diffusion networks: A mixed diffusion pattern model," *Knowledge-Based Systems*, vol. 147, pp. 81 – 93, 2018.

[26] S. Feizi, M. Medard, G. Quon, M. Kellis, and K. Duffy, "Network infusion to infer information sources in networks," *IEEE Transactions on Network Science and Engineering*, pp. 1–1, 2018.

[27] E. Sefer and C. Kingsford, "Convex risk minimization to infer networks from probabilistic diffusion data at multiple scales," in *2015 IEEE 31st International Conference on Data Engineering*, April 2015, pp. 663–674.

[28] D. Wang, W. Zhou, J. X. Zheng, S. Wen, J. Zhang, and Y. Xiang, "Who spread to whom? inferring online social networks with user features," in *2018 IEEE International Conference on Communications (ICC)*, May 2018, pp. 1–6.

[29] S. Shaghaghian and M. Coates, "Bayesian inference of diffusion networks with unknown infection times," in *2016 IEEE Statistical Signal Processing Workshop (SSP)*, June 2016, pp. 1–5.

[30] D. Sheldon, T. Sun, A. Kumar, and T. Dietterich, "Approximate inference in collective graphical models," in *Proceedings of the 30th International Conference on Machine Learning*, vol. 28, no. 3, Atlanta, Georgia, USA, 17–19 Jun 2013, pp. 1004–1012.

[31] S. A. Myers, C. Zhu, and J. Leskovec, "Information diffusion and external influence in networks," in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, 2012, pp. 33–41.

[32] W. Webberley, S. Allen, and R. Whitaker, "Retweeting: A study of message-forwarding in twitter," in *2011 Workshop on Mobile and Online Social Networks*, Sep. 2011, pp. 13–18.

[33] Y. Zhou, B. Xu, J. Xu, L. Yang, C. Li, and B. Xu, "Compositional recurrent neural networks for chinese short text classification," in *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, Oct 2016, pp. 137–144.

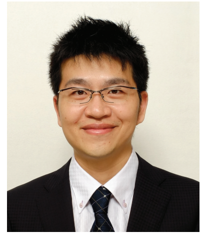
[34] M. Liu, L. Zhang, H. Hu, L. Nie, and J. Dai, "A classification model for semantic entailment recognition with feature combination," *Neurocomputing*, vol. 208, pp. 127–135, 2016, sI: BridgingSemantic.

[35] P. Liu, X. Qiu, and X. Huang, "Recurrent neural network for text classification with multi-task learning," pp. 1–7, 2016. [Online]. Available: arXiv:1605.05101v1

[36] Y. Wang, X. Chen, and J. Li, "A new genetic-based rumor diffusion model for social networks," in *2015 International Conference on Cyber Security of Smart Cities, Industrial Control System and Communications (SSIC)*, Aug 2015, pp. 1–5.

[37] E. Kafeza, A. Kanavos, C. Makris, and P. Vikatos, "Predicting information diffusion patterns in twitter," in *Artificial Intelligence Applications and Innovations*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 79–89.

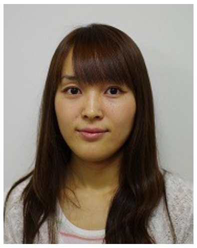
Xiao Yang was born in Chongqing, China. He received B.S. degree in Telecommunication Engineering from the School of Communication and Information Engineering in Chongqing University of Posts and Telecommunications, Chongqing, China in 2017. He is currently purchasing the M.S. degree in Electronics and Communication Engineering at School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China and he is also a visiting student at Muroran Institute of Technology. His research interests include information diffusion, social media, machine learning, natural language processing, etc.



Mianxiong Dong Mianxiong Dong received B.S., M.S. and Ph.D. in Computer Science and Engineering from The University of Aizu, Japan. He became the youngest ever Professor of Muroran Institute of Technology, Japan where he currently serves Advisor to Executive Director, and Vice Director of Office of Institutional Research. He was a JSPS Research Fellow with School of Computer Science and Engineering, The University of Aizu, Japan and was a visiting scholar with BBCR group at University of Waterloo, Canada supported by JSPS Excellent Young Researcher Overseas Visit Program from April 2010 to August 2011. Dr. Dong was selected as a Foreigner Research Fellow (a total of 3 recipients all over Japan) by NEC C&C Foundation in 2011. His research interests include Wireless Networks, Cloud Computing, and Cyber-physical Systems. He has received best paper awards from IEEE HPCC 2008, IEEE ICESS 2008, ICA3PP 2014, GPC 2015, IEEE DASC 2015, IEEE VTC 2016-Fall, FCST 2017, 2017 IET Communications Premium Award and IEEE ComSoc CSIM Best Conference Paper Award 2018. Dr. Dong serves as an Editor for IEEE Transactions on Green Communications and Networking (TGCN), IEEE Communications Surveys and Tutorials, IEEE Network, IEEE Wireless Communications Letters, IEEE Cloud Computing, IEEE Access, as well as a leading guest editor for ACM Transactions on Multimedia Computing, Communications and Applications (TOMM), IEEE Transactions on Emerging Topics in Computing (TETC), IEEE Transactions on Computational Social Systems (TCSS). He has been serving as the Vice Chair of IEEE Communications Society Asia/Pacific Region Information Services Committee and Meetings and Conference Committee, Leading Symposium Chair of IEEE ICC 2019, Student Travel Grants Chair of IEEE GLOBECOM 2019, and Symposium Chair of IEEE GLOBECOM 2016, 2017. He is the recipient of IEEE TCSC Early Career Award 2016, IEEE SCSTC Outstanding Young Researcher Award 2017, The 12th IEEE ComSoc Asia-Pacific Young Researcher Award 2017, Funai Research Award 2018 and NISTEP Researcher 2018 (one of only 11 people in Japan) in recognition of significant contributions in science and technology. He is currently the Member of Board of Governors and Chair of Student Fellowship Committee of IEEE Vehicular Technology Society, and Treasurer of IEEE ComSoc Japan Joint Sections Chapter. He is Clarivate Analytics 2019 Highly Cited Researcher (Web of Science).

Xiuzhen Chen received Ph.D. degree in Xi'an Jiao Tong University, Xi'an, China. She is currently an Associate Professor in Shanghai Jiao Tong University and a Certified Information System Auditor (CISA). She was also a visiting scholar to Télécom Paris Tech.

She has presided and participated in more than 20 national projects, including National Key R&D Program of China, National Natural Science Foundation of China, and the National Key Research and Development Program in Cyberspace Security. Her research interests focus on computer network security, intelligent automobile security and online social network analysis.



Kaoru Ota Kaoru Ota was born in Aizu-Wakamatsu, Japan. She received M.S. degree in Computer Science from Oklahoma State University, USA in 2008, B.S. and Ph.D. degrees in Computer Science and Engineering from The University of Aizu, Japan in 2006, 2012, respectively. She is currently an Associate Professor with Department of Sciences and Informatics, Muroran Institute of Technology, Japan. From March 2010 to March 2011, she was a visiting scholar at University of Waterloo, Canada. Also she was a Japan Society of the Promotion of Science (JSPS) research fellow with Kato-Nishiyama Lab at Graduate School of Information Sciences at Tohoku University, Japan from April 2012 to April 2013. Her research interests include Wireless Networks, Cloud Computing, and Cyber-physical Systems. Dr. Ota has received best paper awards from ICA3PP 2014, GPC 2015, IEEE DASC 2015, IEEE VTC 2016-Fall, FCST 2017, 2017 IET Communications Premium Award and IEEE ComSoc CSIM Best Conference Paper Award 2018. She is an editor of IEEE Transactions on Vehicular Technology (TVT), IEEE Internet of Things Journal, IEEE Communications Letters, Peer-to-Peer Networking and Applications (Springer), Ad Hoc & Sensor Wireless Networks, International Journal of Embedded Systems (Inderscience) and Smart Technologies for Emergency Response & Disaster Management (IGI Global), as well as a guest editor of ACM Transactions on Multimedia Computing, Communications and Applications (leading), IEEE Internet of Things Journal, IEEE Communications Magazine, IEEE Network, IEEE Wireless Communications, IEEE Access, IEICE Transactions on Information and Systems, and Ad Hoc & Sensor Wireless Networks (Old City Publishing). She is the recipient of IEEE TCSC Early Career Award 2017, and The 13th IEEE ComSoc Asia-Pacific Young Researcher Award 2018.She is Clarivate Analytics 2019 Highly Cited Researcher (Web of Science).